ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2024.0490

LEVERAGING DATA-DRIVEN TECHNIQUES FOR EFFICIENT DATA MINING IN CLOUD COMPUTING ENVIRONMENTS

Jaladurgam Venkataramana
Dell EMC Corporation, Bengaluru, India

Abstract

The capacity to efficiently use big data and analytics is becoming a critical differentiator for company growth in today's data-driven environment. Using important trends, obstacles, and best practices as a framework, this article investigates how to promote company growth via the use of big data and analytics. An important issue in cloud computing is deciding on an acceptable amount and location of data. Decisions about resource management are based on data aspects and operations in data-driven infrastructure management (DDIM), a novel solution to this problem. It is critical to have a unified system that can manage various forms of big data and the analysis of that data, as well as common knowledge management functions. The approach stated in this research is DD-DM-CCE, or Data-Driven Methods for Efficient Data Mining in Cloud Computing Environments. Improving data using derived information from maximum frequent correlated pattern mining is the main focus of the work. By considering the centrality factor, the DD-DM-CCE method may help choose the best locations to store data in order to reduce access latency. In order to gain a competitive edge, this study offers a cloud-based conceptual framework that can analyze large data in real time and improve decision making. Efficient big data processing is possible with cloud computing infrastructures that can store and analyze massive amounts of data, as this reduces the upfront cost of the massively parallel computer infrastructure needed for big data analytics. According to simulations run on cloud computing, the DD-DM-CCE approach does better than the status quo regarding hit ratio, effective network utilization, and average response time. According to this study, data mining methods are valuable and successful in predicting how consumers will utilize cloud services.

Keywords:

Data-Driven, Data Mining, Cloud Computing, Big Data

1. INTRODUCTION

Some grid and cloud computing applications confront data-intensive systems, comprising tasks that analyse or transport data on a big scale. Recently there has been interest in cloud computing systems as a potential new standard for data administration. However, the need for simple storage components that are easy to manage on a big scale grows with user demands, data file sizes, and requests to extract more insights from massive data sets [1-2]. Data is one of the most important parts of cloud computing; thus, offering a reliable, efficient, well-maintained platform is crucial. Data availability, fault tolerance, durability, and reliability must be presented at various levels via well-defined data management approaches. In addition, the features mentioned earlier, and reaction time may be improved on a platform that applies data replication, which is significant from the consumer's point of view. Like electricity and water, cloud computing offers various computing services, such as utility [3-4].

This paper presented data mining as a method for automatically obtaining valuable data and insights derived from large datasets—including correlations, patterns, trends, and structures. Learning the application's domain, gathering data, cleaning and preparing it, choosing a data mining technique, and obtaining an assessment are all parts of a standard knowledge discovery process [5]. The three primary data mining methods that rely on distributed settings are clustering, association analysis, and classification. Grid and cloud computing are all about distributed systems and protocols, while data mining is all about combining driven and scheduling techniques to find insights into data. These days, cloud computing and data mining are working together in a new way, thanks to the cloud's proliferation and the growing complexity of data mining applications [6-7]. Distributed data mining applications may take advantage of the robust computing resources offered by cloud environments. However, a new class of data mining applications is taking shape, and one of them is mining data stored on the cloud. Monitoring cloud access patterns using data mining methods to extract new, relevant insights and improve cloud performance is an intriguing challenge [8-9]. In most cases, you'll find that some data files are associated with the same individuals, apps, or occupations. Coordination of file requests is joint in many real-world data-intensive domains, including high-energy physics. This paper may put highly correlated files in the exact driven group to improve the data-driven technique. One way to decrease dispersed transactions needing access to several sites is to keep associated files in the same or neighboring locations. Data mining is the foundation of our proposed data [10-11].

Data-Driven Infrastructure Management (DDIM) solutions that are new to this field have been created and shown in recent publications. In order to identify problems and mistakes in physical systems, there are experimental DDIM solutions that analyze data streams from infrastructure [12-13]. In some instances, irregularities in cloud infrastructure functioning have been detected using more complex data mining approaches, such as machine learning. Cloud infrastructure performance management based on data and handling resources based on data are supporting developments. But a more comprehensive Data-Driven Infrastructure Management strategy is needed, and current solutions only deal with optimization and resilience on an individual level [14-15].

They are driven by (DD-DM-CCE), which uses access history to establish file correlations. Good performance benefits are possible with an efficient data-driven placement approach. Here, DD-DM-CCE chooses the optimal location based on its centrality and the number of visitors. Last, DD-DM-CCE significantly improves storage resource consumption efficiency using a paradigm based on driven deletion instead of superfluous replication. If DD-DM-CCE determines that a data file will not be required soon, it will delete it. The DD-DM-CCE is modelled for assessment purposes in cloud computing. The experimental findings demonstrate that DD-DM-CCE may improve cloud network utilization, decrease access time, and increase hit ratio and total communication volume.

Main Contribution,

- This article presents the effort to develop data-driven applications in a serveries cloud computing environment using a data-driven approach.
- This paper presents a data model for the cloud computing environment to assist the CSB in supporting virtual resource allocation, control, and management between cloud users and CSPs.
- Additionally, the data flow model clarifies the stages and levels of data from both on-premises and cloud infrastructure, allowing for analyzing vulnerabilities and threats at each stage to establish appropriate controls for ensuring overall security.

The remaining sections of the document are organized as follows: The work's literature review is detailed in Section 2. The idea behind the suggested model is discussed in Section 3. The performance assessment of the proposed model is discussed in Section 4. Section 5 provides suggestions for further study and serves as the paper's last section.

2. RELATED SURVEY

To enhance the detection system's accuracy in a cloud computing environment, Aws Naser Jaber et al. [16] proposed a new intrusion detection system that integrates a fuzzy clustering method (FCM) with support vector machines (SVM). The paper implemented the suggested approach and evaluated it against current methods. Experiments are conducted using the NSL-KDD dataset. Using this new hybrid mechanism (FCM-SVM), tests of performance and comparisons with current methods show that the suggested system does better in finding things correctly and reducing the number of false alarms.

Jin Gao et al. [17] proposed DMAD-DVDMR, a data mining method for finding anomalies in the cloud, as an alternative to traditional slow-train algorithms, which need to work better at predicting the future and give unstable detection results. Data is preprocessed using a dimensionality reduction methodology. Deep variational learning is used in this model to reduce data dimensionality and computation strain while providing complete data. Second, the Hadoop Distributed File System (HDFS) uses MapReduce to process data in parallel; therefore, each data point's k-distance and LOF value can only be calculated inside its block. Experimental data shows that the DMAD-DVDMR approach is successful and scalable.

Yuan Jiugen et al. [18] examine the meaning of big data mining (BDM) and suggest a large data mining system frame, including cloud computing and mining services, all while contrasting the topic with conventional data mining. This article examines its internal process to help readers understand and use big data mining, using the multipurpose Hadoop platform as an example. It also discusses the platform's benefits and drawbacks.

Some notable quantum algorithms by Sudharson K et al. [19] include QPCA, Grover's search approach, and quantum support vector machines (QSVM). Quantum computing is a promising technique for data mining in cloud settings due to its parallel processing capabilities. The paper also compares various approaches and examines data mining challenges using quantum computing, including the need for specialized expertise, concerns about scalability, and hardware limitations. Results show that quantum computing works well for cloud-based data mining on a massive scale, and they also point the way for further studies in the field.

To train an intrusion detection system (IDS) to recognize harmful network traffic, Amirah Alshammari et al. [20] suggested using machine learning (ML), a popular approach. The reliability of the training dataset has a significant impact on the detection accuracy of ML models. To find malicious patterns in network traffic, this research suggests an intrusion detection system technique that uses ML models. The dataset that this detection method uses includes both malicious and benign traffic. Obtaining the variables needed to train the ML model to differentiate between normal and abnormal data presents a substantial challenge to the research.

3. PROPOSED METHOD

Internet and mobile app users in today's lightning-fast world want their data to be immediately updated in the cloud anytime there's a modification or update to the database, and they want it shown on their screens without reloading the client app. The Fig.1 demonstrates the architecture for data-driven applications that use real-time data streaming and event-driven architecture to accomplish these criteria.

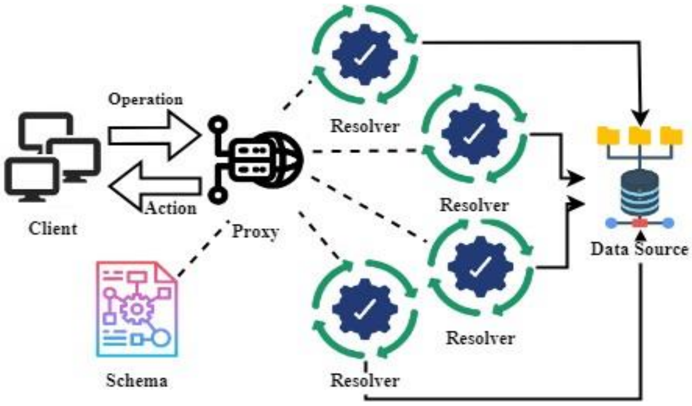


Fig.1. Data-driven serveries cloud computing architecture

The use of the cloud for storing, processing, and distributing data has grown in popularity and appeal. Without the huge upfront costs sometimes associated with deploying conventional data centers, it offers elastic computing and data storage capabilities on demand. Therefore, big data analytics and storage have necessitated infrastructures capable of handling massive amounts of data, and the rise of cloud computing has provided just that. To that end, we're working to build a smart cloud-based computing framework that can automate and improve the execution of a big data-based CC by efficiently analyzing and sharing big data collected from various sources in real-time. Therefore, businesses may reap the benefits of cloud computing, which allows for the on-demand delivery of big data analytics as a cloud service and the rapid adjustment of IT resources inside to suit fluctuating needs.

- **Client**: Any program or user engaging with a cloud service is referred to as a client. Clients start data processing, querying, and data mining tasks. Operations and actions sent to the system by clients initiate data mining activities.

- **Operation and Action**: Here, the customer wants to carry out specific actions or give them instructions. They may include data-driven activities such as conducting analytics or queries. They aim to represent the user's intent so that the following architectural components may understand it.
- **Proxy**: The proxy mediates communication between the client and the many components processing the data. Its job is to process and direct the customer's requests. When a client initiates an operation, the proxy takes it in and passes it on to the relevant resolvers. Additionally, it may validate or process the requests first.
- **Schema**: The schema represents the data model or structure being processed or queried. In the cloud, it specifies how data is structured, saved, and retrieved. The schema guides data retrieval and processing by resolvers and checks whether the data operations are in sync with the data structure.
- **Resolvers**: Components that handle the client's requests are called resolvers. Some resolvers may be more adept at certain operations than others or collaborate to complete tough jobs. When the proxy sends a request, the resolvers use the schema to determine how to handle it. They extract helpful information from the data source via data mining, processing, and analytics.
- **Data Source**: Data is stored in a system known as the data source. It may be cloud storage, such as a database or data warehouse. To carry out the desired activities, the resolvers rely on the data provided by the data source. All data handled by this architecture originates from it.

Effective data mining in a cloud setting makes use of data-driven approaches. Resolvers receive operations from the client and process them according to a preset schema after passing them via a proxy. At last, the client's request is fulfilled when the data is obtained from a data source. This approach is perfect for cloud-based situations where data mining is essential since it guarantees scalability, flexibility, and fast processing of processes that are heavy on data.

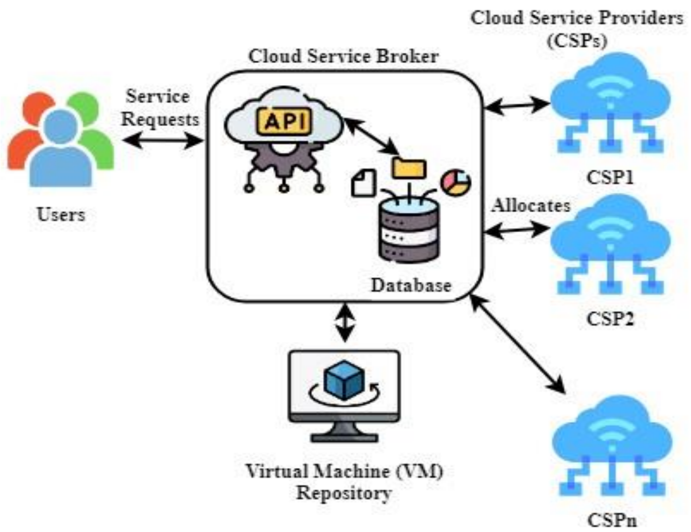


Fig.2. Cloud Services Brokerage

The Fig.2 shows that cloud computing is this study's subject. The components that make up this ecosystem include users, VMRs, CSPs, PMs, and DCs. A cloud services broker is a third party that mediates between CSPs and their consumers that use cloud services. Cloud service brokers act as intermediaries between their customers and the many cloud service providers, negotiating the rental of different cloud resources on their behalf. A cloud service broker's responsibility includes (i) figuring out where in the cloud a virtual infrastructure's resources should be most effectively deployed, (ii) keeping an eye on and managing those resources, and (iii) integrating several cloud services into one customized offering for clients. As stated by DD-DM-CCE, the following abilities and qualifications are necessary for a cloud service broker to provide brokerage services.

Ensure that the CSPs' service meets the cloud user's needs.

- Get into service level agreement (SLA) negotiations with cloud service providers and consumers.
- Deploy CSP services to cloud users effectively.
- Act in response to SLA breaches and conduct regular performance reviews of these SLAs.
- Preserve the privacy of customers' information and the security of CSP's service.
- Ensure that all CSPs apply access control decisions consistently.
- Draw a secure picture of the CSPs' access and identity management systems.

Nevertheless, a CSB can only work correctly with a database management system.

The Fig.1 shows the components of a cloud computing ecosystem, including users, data centers (DC), virtual machines (VM), cloud service providers (CSP), physical machines (PM), and cloud service brokers (CSB). They assess a collection of CSPs following Eq.(1):

$$CSP=\{csp_1,csp_2,\ldots csp_o\} \tag{1}$$

where o is the total number of CSPs under the CSB's management. Every single CSP is made up of DC, following Eq.(2):

$$\text{DC-}\{dc_1,dc_2,\ldots.,dc_n\} \tag{2}$$

where n is the number of DCs that are contained within a CSP, and each DC is comprised of a member of the PMs, following Eq.(3):

$$\text{PM}=\{pm_1,pm_2,\ldots,pm_r\} \tag{3}$$

where r denotes the number of PMs in a DC, the set indicates that each PM is in charge of hosting u virtual machines following Eq.(4):

$$\text{VM}=\{vm_1,vm_2,\ldots,vm_u\} \tag{4}$$

They believe that CSPs should be given the responsibility of assigning L tasks. A total of x_k virtual machines are required for every task l. The notation is used to indicate the content of the CSB allocation by following Eq.(5),

$$l\rightarrow vm_{xk} \tag{5}$$

According to Eq.(6), either x_k=1 or $x_k \leq u$. One p_m is occupied by every v_m,

$$v_m\rightarrow p_m \tag{6}$$

One data center d_c is responsible for hosting each pm, following Eq.(7):

$$p_m \rightarrow d_c \tag{7}$$

Eq.(8) shows that every data centre is owned by a single cloud service provider (CSP).

$$d_c \rightarrow csp \tag{8}$$

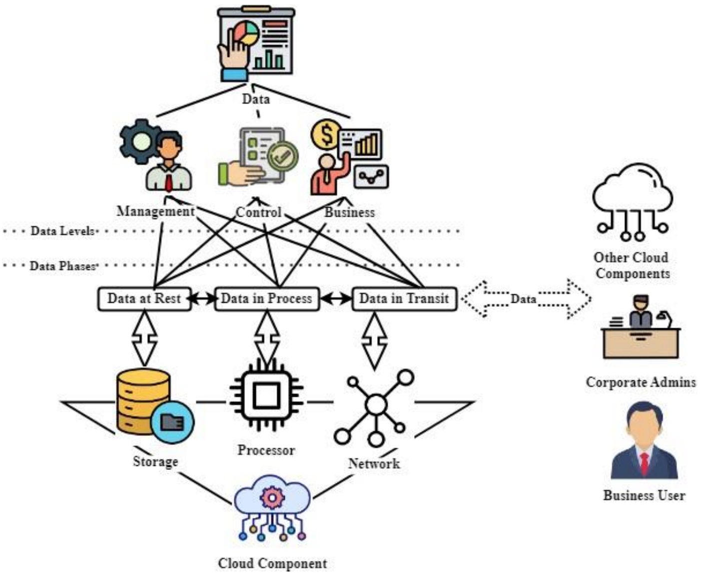


Fig.3. Abstraction of Data Level

The data in the cloud differs from data in conventional infrastructure because it is transferred and stored in the cloud service provider's (CSP) infrastructure. Understanding different forms of data is crucial throughout the DD-DM-CCE's lifespan. DD-DM-CCE identifies three abstracted levels for data classification: management, control, and business. Each level considers three data states: at rest, in process, and in transit.

- Data created by the organization at the management level (mD) pertains to the administration of cloud components, including the cloud admin interface and cloud computing resources. The information may include methods for authentication or access control.
- Data at the control level (cD) pertains to information that facilitates technological operations, including, but not limited to, application queries, configuration changes, logs, action events, data backup and replication, application programmable interface (API) calls, routing information, and application data. Control data might be only operational data, or it can be more valuable to companies, including consumer information requests for APIs or authentication of identities. Nonetheless, information is eager to be shared across on-premises and remote servers.
- Any data about business services that users, such as email, create when using organizational cloud services is known as business data level (bD).

Particularly in situations where business and operational data are closely connected, companies must safeguard any data that might potentially disclose any possible danger. Data leakage may occur if there is a vulnerability in the application programming interface (API) connection between cloud components. In addition, hacks that take control of an account might have a significant effect since cloud services are not encrypted. In addition to the data at the control and management levels, the cloud platform also needs data at the business level. This data is stored on the system and includes information shared between cloud service users and providers. Even though it is still susceptible to threats, this data must be protected inside the cloud platform for organizations that use a cloud-based system. Depending on the security at each management and control level, this information may be protected from the general public or made available.

- There is a term for when data is stored on a remote server or the infrastructure of a cloud service provider (CSP): "at rest" (Dr). The data is also kept for long-term uses as it reaches the later stages of the cycles in the cloud. The system's security controls may compromise, alter, or delete it while it is still out of sight.
- A phrase used to represent data that is now inactive but has been available for a brief time by a user or service is "Data in Process" (Dp). The execution of an application may make use of this data at various points. Data like this isn't inherently suspicious, but a malicious service, program, or hardware may change or leak it. The process of data being transported to other users or systems is referred to as "Data in Transit" (Dt). Cloud computing platforms pose a significant threat to data in transit when it travels across unsecured networks or to an API that facilitates application-to-application communication. Administrative management IDs and other operational and commercial data are shared online in a cloud environment. Data in transit is vulnerable to various attacks, including Man-in-the-Middle attacks. Even if data security is receiving more and more attention from legislators, this suggests that the data phase should still be protected. This is because a data breach may have far-reaching consequences for a business, including exposing sensitive information, which can harm its reputation or result in financial penalties. A security breach may have occurred if unauthorized individuals accessed system resources and accessed data via Advanced Persistence Threat or malicious hardware components. To launch a data leak assault against Amazon Web Services (AWS), AWS implanted a rogue microchip into Supermicro computers.
- The term "Data in Transit" (Dt) describes the process by which data are transferred to other users or systems. Data in transit is highly susceptible to compromise by cloud computing systems when flowing across insecure networks or to an application programming interface (API) that enables applications to communicate with one another. Within a cloud environment, business and operational data, such as administrative management identities, are exchanged over the Internet. Several other types of assaults may be launched against data in transit, including Man-in-the-Middle attacks. According to this, the data phase needs to be safeguarded despite the increasing legislative attention on data security. This is because a data breach may impact various levels of an organization, such as the disclosure of sensitive material, which can result in reputational damage or financial fines.

The data levels and stages present in every cloud system are shown in Fig.3. The graphic displays data originating from any cloud user at three different levels of abstraction. There is a connection between each level and the user type (actor). Potential players in this situation include operators, corporate users, or even other systems. Any data level might be present during data storage, processing, or transmission. In contrast, the three main parts of cloud infrastructure are the CPU, memory, and the network, responsible for input/output. Cloud infrastructure is linked to data stages.

4. RESULT AND DISCUSSION

Efficiency and accuracy in data mining were significantly enhanced with the use of data-driven methodologies in cloud computing settings. These advancements result from combining cloud computing with machine learning algorithms, which allow for scalable, real-time data processing. The system's capability to adapt to diverse kinds of data and workloads is shown in the discussion, which emphasizes its adaptability in different cloud settings. Protecting sensitive data is essential in cloud computing, but the architecture's built-in security mechanisms take it a step further. These results indicate that the DD-DM-CCE design facilitates future developments in cloud-based analytics by optimizing data mining operations and providing a safe and adaptable framework.

4.1 DATASET

The growing need for actionable insights, improved decision-making, and digitized data drives the Indian data analytics industry. With IoT and cloud computing, organizations create massive volumes of data, which drive market development. Additionally, the requirement to get relevant insights from this data has increased, driving data analytics solution development. Government efforts like "Digital India" also drive data analytics industry development. These programs foster technology improvements and digital literacy, fostering data analytics growth. Thus, local and foreign vendors actively engage in the India data analytics market, providing novel solutions to match organizations' changing demands to maximize data value. The India data analytics market is expected to increase, allowing organizations to use data-driven insights for sustainable growth and competitiveness [21].

4.1.1 Accuracy:

To determine if our method is effective in data segmentation and classification, this paper put it through its paces, utilizing the challenge evaluation criteria. Here are the defined success criteria:

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \quad (9)$$

The Eq.(9), which is the ratio of the number of assessments to the number of correctly predicted values (where TN stands for true negative, meaning that a negative observation is correctly predicted as unfavourable), can be used to Fig.4 out how well a model finds correlations and patterns in datasets.

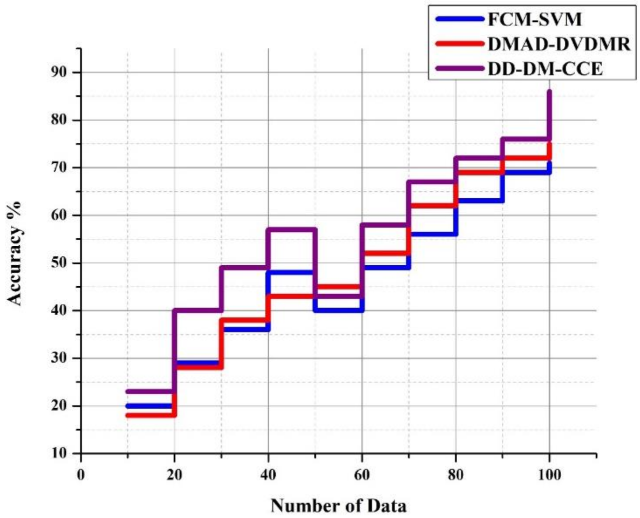
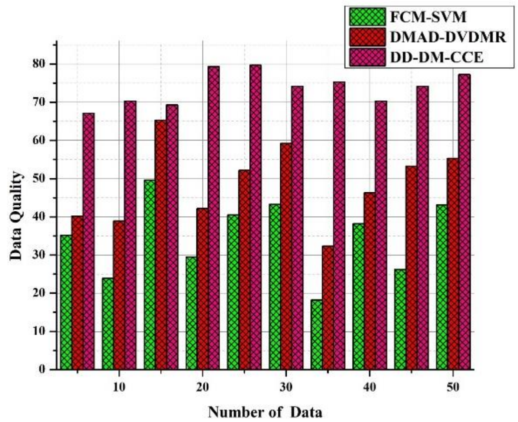


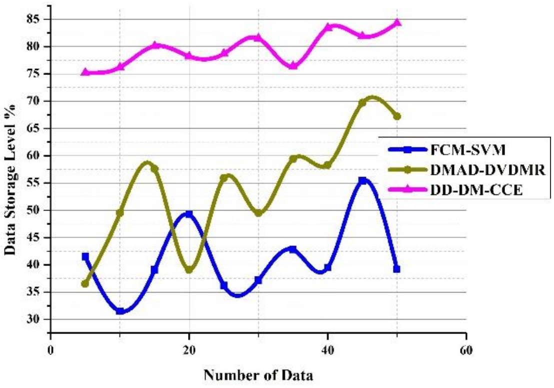
Fig.4. Accuracy

The Fig.4 displays the Accuracy curves of the refined DD-DM-CCE composites, as determined by Eq.(9). By incorporating state-of-the-art data mining techniques into the DD-DM-CCE architecture, data mining tasks were significantly improved. Compared to more conventional data mining methods, the architecture significantly enhanced pattern and anomaly detection accuracy in complicated and massive datasets. Healthcare analytics, financial projections, and fraud detection need high reliability; therefore, this enhancement is crucial. The system can continually improve models with fresh data inputs while effectively managing massive volumes of data in real-time, improving accuracy. The DD-DM-CCE architecture provides reliable and timely insights for cloud-based data mining that requires precision.

4.1.2 Data Quality Level vs Data Storage:



(a)

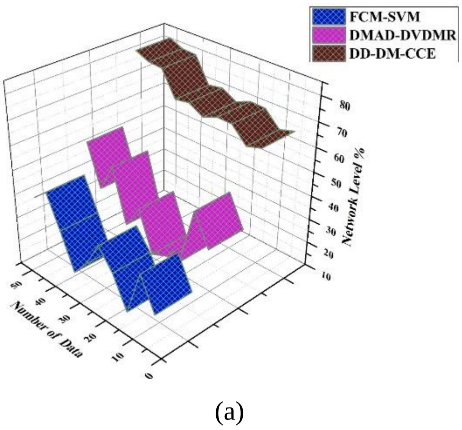


(b)

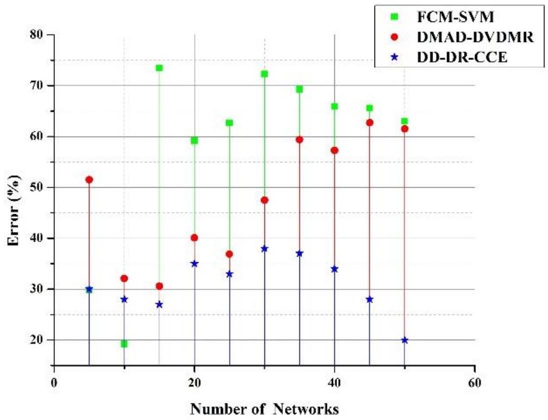
Fig.5. Data Quality Level vs Data Storage

The Fig.5 (a) displays the Data Quality curves of the refined DD-DM-CCE composites, as determined by Eq.(s 1 to 4. Data quality impacts research efficacy in the DD-DM-CCE design. Quality data should be accurate, dependable, and valuable. This study showed that data mining accuracy depended on data quality. Data was cleaned, normalized, and validated to eliminate noise, inconsistencies, and outliers. Structured and clean input data let the architecture apply deep-learning approaches to find significant trends and patterns. However, correct conclusions, analyses, or projections may lead to better data quality. Effective cloud-based data mining requires high-quality data, which the DD-DM-CCE paradigm prioritizes. The Fig.5(b) shows that the graph compares the data storage levels for three methods, FCM-SVM, DMAD-DVDMR, and DD-DM-CCE, across varied data counts. The DD-DM-CCE method maintains most data storage levels between 75% and 80%. Storage levels vary between 45 and 60% when using the DMAD-DVDMR technique, which is more variable. With storage levels ranging from 30% to 55%, the FCM-SVM approach is both the lowest and most unpredictable. Compared to other methods, DD-DM-CCE's consistent performance shows that it manages data storage levels more efficiently.

4.1.3 Network Level vs Network Error:



(a)

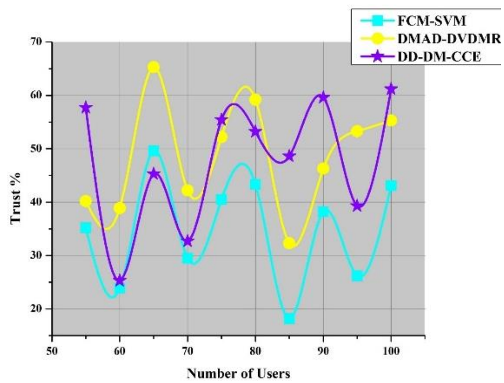


(b)

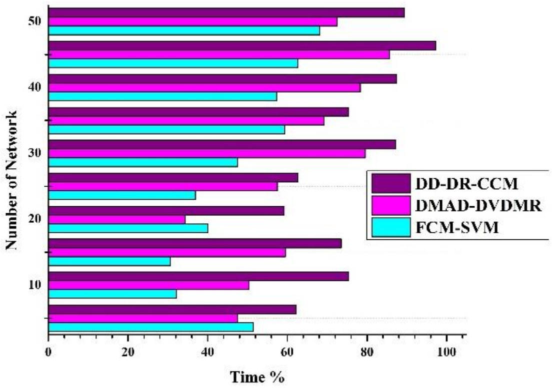
Fig.6. Network Level vs Network Error

The Fig.6 (a) displays the Network level curves of the refined DD-DM-CCE composites, as determined by Eq.(s 5 to 7. Since cloud computing data mining speed, reliability, and efficiency depend on the network, the DD-DM-CCE architecture examines it. A fast, dependable network is required to transfer enormous volumes of data between cloud resources and data processing units. According to studies, network bandwidth and latency were essential for data transmission and real-time processing. High network bandwidth minimized design constraints and expedited data processing. Time-sensitive applications require low latency to reduce data source, storage, and processing unit connection delays. Network scalability lets the design handle greater workloads without affecting performance. Finally, data-driven cloud-based data mining methods need a well-optimized network for fast and accurate analysis. Well-optimized networks increase data security and user connectivity. The Fig.6(b) this graph compares the error rates (%) for three different methods—FCM-SVM, DMAD-DVDMR, and DD-DR-CCE—at different network densities. The DD-DR-CCE algorithm maintains an error rate range of 20% to 30% as the number of networks rises and consistently obtains the lowest rate. The DMAD-DVDMR method has error rates that are both larger and more variable, often falling within the 30% to 50% range. With large swings that may go above 70%, the FCM-SVM algorithm has the worst error rates. Based on these results, the DD-DR-CCE algorithm is the best option compared to the other two algorithms to reduce error rates, particularly as complexity (number of networks) rises.

4.1.4 User's Trust Level vs Network Time:



(a)



(b)

Fig.7 (a) and (b): User's Trust Level vs Network Time

The Fig.7(a) displays the Trust curves of the refined DD-DM-CCE composites, as determined by Eq.(8. Users' comfort and confidence in data mining in the cloud depend on how much they trust the DD-DM-CCE architecture. Methods build trust. Users trust this research because of the architecture's timeliness and accuracy. User data was encrypted and restricted, so they felt secure. Open data processing and analysis allowed customers to understand and confirm results. Users felt safe because access restrictions and encryption protected their data. CE architecture promotes user confidence by considering these. The Fig.7(b) this graph compares FCM-SVM, DMAD-DVDMR, and DD-DR-CCE task completion times across various network counts. Even as the number of networks expands, the DD-DR-CCE method remains the fastest with stable processing times. The DMAD-DVDMR approach is slower, and processing times rise with network count. FCM-SVM is the slowest algorithm, particularly when processing several networks. This comparison shows that the DD-DR-CCE approach is the fastest for complex projects with many networks. The DD-DR-CCE algorithm improves accuracy and dependability.

5. CONCLUSION

The study concludes that data-driven approaches improve the cloud computing DD-DM-CCE architecture. The results suggest that this strategy enhances data mining accuracy and efficiency and provides a secure, scalable framework for big and complex datasets. The architecture's faster processing and accurate pattern detection enhance healthcare and finance data analytics. It is possible to increase the performance and practicability of the DD-DM-CCE architecture using various analysis methods. Deep learning or reinforcement learning may assist in resolving more sophisticated data mining issues. Integrating design with edge computing can enhance data processing near the data source, decreasing latency and bandwidth use. Hybrid cloud solutions, which combine the most advantageous aspects of public and private clouds, are yet another fascinating new sector that has the potential to revolutionize data management. Future architectural development may concentrate on managing unstructured data, including text, photos, and video. In data-driven decision-making, this kind of data is becoming more critical. Improving the architecture's compatibility with other cloud-based systems would make it more adaptable and accessible to incorporate into existing activities. It is necessary to strengthen architecture cybersecurity in order to secure data and battle newly developing threats. In conclusion, the DD-DM-CCE architecture is an effective instrument for cloud data mining, but continuous improvement and innovation are required to stay up with the latest technological developments. In a data-driven environment, the architecture will be able to remain relevant and provide value by addressing these potential problems and opportunities.

REFERENCES

[1] D.K. Sharma, A. Dharmaraj, A. Al Ayub Ahmed, K. Suresh Kumar, K. Phasinam and M. Naved, "A Study on the Relationship Between Cloud Computing and Data Mining in Business Organizations", *Proceedings of International Conference in Mechanical and Energy Technology*, pp. 91-99, 2022.

[2] M. Shengdong, X. Zhengxian and T. Yixiang, "Intelligent Traffic Control System based on Cloud Computing and Big Data Mining", *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 12, pp. 6583-6592, 2019.

[3] J. Wu, N. Mu, X. Lei, J. Le, D. Zhang and X. Liao, "SecEDMO: Enabling Efficient Data Mining with Strong Privacy Protection in Cloud Computing", *IEEE Transactions on Cloud Computing*, Vol. 10, No. 1, pp. 691-705, 2019.

[4] S. Dhanasekaran, K. Rajput, M. Aeri, R.P. Shukla and S.K. Singh, "Utilizing Cloud Computing for Distributed Training of Deep Learning Models", *Proceedings of International Conference on Data Science and Information System*, pp. 1-6, 2024.

[5] V.A.K. Gorantla, V. Gude, S.K. Sriramulugari and P. Yadav, "Utilizing Hybrid Cloud Strategies to Enhance Data Storage and Security in E-Commerce Applications", *Proceedings of International Conference on Disruptive Technologies*, pp. 494-499, 2024.

[6] S.K. Sriramulugari, V.A.K. Gorantla, V. Gude and K. Gupta, "Exploring Mobility and Scalability of Cloud Computing Servers using Logical Regression Framework", *Proceedings of International Conference on Disruptive Technologies*, pp. 488-493, 2024.

[7] H. Osman, M.A. Maarof and M.M. Siraj, "Hybrid Solution for Privacy-Preserving Data Mining on the Cloud Computing", Proceedings of International Conference on Emerging Trends in Intelligent Computing and Informatics: Data Science, Intelligent Information Systems and Smart Computing, Vol. 4, pp. 748-758, 2020.

[8] P. Zheng, X. Xu and C.H. Chen, "A Data-Driven Cyber-Physical Approach for Personalised Smart, Connected Product Co-Development in a Cloud-based Environment", *Journal of Intelligent Manufacturing*, Vol. 31, No. 1, pp. 3-18, 2020.

[9] J. Yu, "Research on the Improvement of Data Mining Algorithm Based on "Internet+" System in Cloud Computing", *Proceedings of International Conference on Industrial Application of Artificial Intelligence*, pp. 509-513, 2020.

[10] Z. Wang, N. Wang, X. Su and S. Ge, "An Empirical Study on Business Analytics Affordances Enhancing the

Management of Cloud Computing Data Security", *International Journal of Information Management*, Vol. 50, 387-394, 2020.

[11] B. Lin, F. Zhu, J. Zhang, J. Chen, X. Chen, N.N. Xiong and J.L. Mauri, "A Time-Driven Data Placement Strategy for a Scientific Workflow Combining Edge Computing and Cloud Computing", *Transactions on Industrial Informatics*, Vol. 15, No. 7, pp. 4254-4265, 2019.

[12] R. Krishankumar, R. Sivagami, A. Saha, P. Rani, K. Arun and K.S. Ravichandran, "Cloud Vendor Selection for the Healthcare Industry using a Big Data-Driven Decision Model with Probabilistic Linguistic Information", *Applied Intelligence*, Vol. 52, No. 12, pp. 13497-13519, 2022.

[13] H.B. Barua and K.C. Mondal, "A Comprehensive Survey on Cloud Data Mining Frameworks and Algorithms", *ACM Computing Surveys*, Vol. 52, No. 5, pp. 1-62, 2019.

[14] M. Rambabu, S. Gupta and R.S. Singh, "Data Mining in Cloud Computing: Survey", *Proceedings of International Conference* on *Innovations in Computational Intelligence and Computer Vision*, pp. 48-56, 2020.

[15] N. Dai, "Retracted Article: Application and Functional Simulation of Data Mining Technology in Hadoop Cloud Platform based on Improved Algorithm", *Soft Computing*, Vol. 27, No. 12, pp. 8381-8389, 2023.

[16] A.N. Jaber and S.U. Rehman, "FCM-SVM based Intrusion Detection System for Cloud Computing Environment", *Cluster Computing*, Vol. 23, No. 4, pp. 3221-3231, 2020.

[17] A. Alshammari and A. Aldribi, "Apply Machine Learning Techniques to Detect Malicious Network Traffic in Cloud Computing", *Journal of Big Data*, Vol. 8, No. 1, pp. 1-25, 2021.

[18] J. Gao, J. Liu, S. Guo, Q. Zhang and X. Wang, "A Data Mining Method using Deep Learning for Anomaly Detection in Cloud Computing Environment", *Mathematical Problems in Engineering*, pp. 1-11, 2020.

[19] Y. Jiugen and X. Ruonan, "Cloud Computing-based Big Data Mining Connotation and Solution", *Proceedings of International Conference on Computer Science and Education*, pp. 245-248, 2020.

[20] K. Sudharson and B. Alekhya, "A Comparative Analysis of Quantum-based Approaches for Scalable and Efficient Data Mining in Cloud Environments", *Quantum Information* and *Computation*, Vol. 23, No. 9, pp. 783-813, 2023.

[21] Leveraging Data Driven Techniques, Available at https://datasetsearch.research.google.com/search?src=0&query=Leveraging%20Data-Driven%20Techniques%20for%20Efficient%20Data%20Mining%20in%20Cloud%20Computing%20Environments&docid=L2cvMTF2ejQxbGswYg%3D%3D, Accessed in 2023.